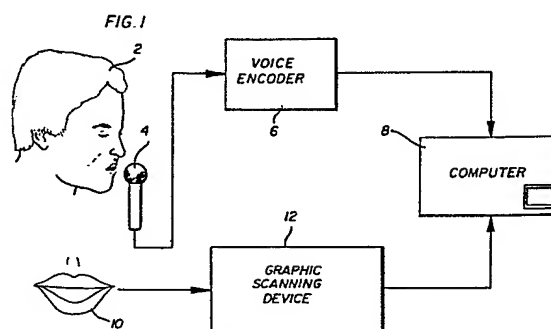(12) # EUROPEAN PATENT APPLICATION

(71) Applicant: Bloomstein, Richard Welcher
1443 Cavell
Highland Park Illinois 60035(US)

(72) Inventor: Bloomstein, Richard Welcher
1443 Cavell
Highland Park Illinois 60035(US)

(74) Representative: Evershed, Michael et al,
MARKS & CLERK Alpha Tower A.T.V. Centre
Birmingham B1 ITT(GB)

(54) Apparatus and method for creating visual images of lip movements.

(57) An apparatus and method that creates visual images of lip movements on film, video tape, or other recorded media. Speech sounds are analyzed, digitally encoded (6) and transmitted to a data memory device (8). Stored within the data memory device (8) is a program for producing output data that creates visual images of lip movements (16) corresponding to the speech sounds. Under control of the data for the speech sounds, the graphical output from the data memory device (8) is sent to a graphic output device and related display equipment (14) to produce the graphical display (16). This display may be combined with the speech sounds so that the resultant audio-visual composite, such as a film strip, contains lip movements corresponding to the speech sounds.

FIG.1

EP 0 056 507 A1

"APPARATUS AND METHOD FOR CREATING
VISUAL IMAGES OF LIP MOVEMENTS"

This invention relates to an apparatus and a method
for creating visual images responsive to analyzed speech
data so as to produce a graphical representation of known
type such as the lip movements corresponding to the speech
sounds. The invention is particularly suitable for
creating visual images of lip movements in films, video
tapes, and on other recorded media.

In the production of many types of audio-visual
media the speech sounds and the visual images are recorded
simultaneously. For example, in the making of motion
pictures or like audio visual recordings, the voice of the
actor is recorded on the sound track at the same time that
the actor is emitting speech sounds. Where the film is
intended to be played as originally produced, the speech
sounds of the sound track correspond to the lip movements
emitted. However, it frequently happens that the audio
portion or sound track is to be in a language other than
the original one spoken by the actor. Under such
circumstances a new sound track in another language is
"dubbed in". When this is done the speech sounds do not
correspond to the lip movements, resulting in an audio-
visual presentation that looks unreal or inferior.

In animated cartoons it is also a problem to provide lip
movements which correspond to the speech sound. This may
be done, however, by utilizing individual art work or
drawings for the lip movements, sometimes as many as
5   several per second. Because of the necessity of making
numerous drawings by hand or other laborious art tech-
niques, the cost of animated cartoons tends to be expensive.

10

In accordance with this invention predetermined visual
images such as lip movements are graphically created to
correspond with speech sounds so that when the visual
images and speech sounds are recorded on film, video tape
15   or other media, the presentation (listening and viewing)
will tend to be more real and interesting. The method and
apparatus of the invention creates graphical images with
a minimum of human effort by the unique utilization of
computerized graphic techniques.
20

In further accordance with this invention there is pro-
vided a coded representation of speech sounds. These
speech sounds may be associated with lip movements from
visual information, which lip movements do not correspond
25   to the speech sound as in a "dubbed in" sound track of a
motion picture. The coded representations of the speech
sounds are transmitted to a data memory device (e.g. com-
puter) for storage therein. There is also stored in the
data memory device coded data for creating a predetermined
30   graphical representation (e.g. lip movements) corresponding
to the speech sounds. This coded data or software is
intended to respond to the coded data representing the
speech sound so that the coded speech sounds can instruct
the computer to send out graphical signals of new lip
35   movements or other graphical representation corresponding
to the speech sounds. The new lip movement data is thus
transmitted to a graphic output device of known type from

which a suitable graphic display may be created. This
graphic display may be, for example, a video display or
a film frame. The audio or speech portions may be com-
bined in correlation with the graphical display.

5

When a motion picture film having a "dubbed in" sound
track in which lip movements are not in correspondence
with the speech sounds, the encoding of the lip movements
may be done on a frame-by-frame basis. Thus, the coded
10   speech sounds may be extracted from the sound track,
frame-by-frame, and sent to the computer. Likewise, the
computer may receive information as to the mouth position
on each frame as well as information relating to the mouth
shape of the actor. The entire film may be optically
15   scanned on a frame-by-frame basis so that each frame with
mouth location and mouth configuration data may be stored
in the computer in digital form along with data in digital
form as to the analyzed speech sounds. When the informa-
tion is sent out from the computer to the graphical output
20   device, the data for the speech sounds causes the computer
to send out the proper graphical output signals to the
graphical output device corresponding to the particular
speech sounds on the sound track of the film. Thus, the
film is reconstructed to the extent necessary to change
25   the mouth shape and lip movement or configuration to cor-
respond with the speech sound.

Typical apparatus of the present invention for creating
visual images of lip movements comprises means such as a
30   speech analyzer for providing a coded representation of
speech sounds, a data memory device, means for transmit-
ting said coded representation to said data memory device
for storage therein, said data memory device having stored
therein coded data for creating a graphical representation
35   of lip movements corresponding to the speech sounds, a
graphical output device, and means for transmitting to
said graphical output device and under control of the

coded speech representation data for the lip movements
corresponding to the speech sounds.

The apparatus further includes means for extracting the coded
speech sounds from a series of frames of audio and visual
information, means for transmitting said coded speech
sounds to said data memory device on a frame-by-frame basis
for storage therein on that basis. Means are also provided
for optically scanning a series of frames in sequence to
provide an encoded graphical image of the visual information.
Means are provided for transmitting the encoded graphical
image data to the computer or data memory device. Means
are also provided for transmitting from the data memory
device to the graphical output device the visual information
data plus the data for the new lip movements which, in a
corrected film, replace the old lip movements on the film.

This invention will now be described in more detail by way
of example with reference to the drawings in which:-

Fig. 1 is a diagram showing the arrangement for storing of
phoneme and graphic codes and forming part of the present
invention;

Fig. 2 is a diagram showing the manner of using the codes
to display visual images;

Fig. 3 is a modified form of the invention showing the
encoding of lip movement corrections; and

Fig. 4 is a diagram showing an arrangement for graphically
displaying the lip movement corrections encoded by the
arrangement of Fig. 3.

Referring now in more detail to the drawing, and particul-
arly to Fig. 1, there is shown an arrangement for storing

phoneme and graphic codes into a digital electronic computer. One set of codes represents the spoken phoneme of a language (e.g. the English language). The other codes are graphic codes representing visual images of lips of
5   various mouth types such as male, female, cartoon animal, etc. together with orientations of the mouth such as front view, three quarter view, side view, etc.

More particularly, a person such as an actor pronounces
10 phoneme into a voice encoder. The voice encoder translates the phoneme into a digital electronic phoneme code which is transmitted to a digital computer and stored in its memory. A phoneme for entire language may be thus be digitally coded. In conjunction with the phoneme code an artist may
15 draw one or more mouth shapes. A programmer or electronic graphic scanning device encodes the artist's drawing into graphic code, which code is also sent for storage into the electronic digital computer. The foregoing is repeated until a complete set of phoneme and graphic codes are
20 stored in digital form representing the basic phoneme code, standard mouth types and orientations, etc. as heretofore stated.

A phoneme code is a representation of intensities of sound
25 over a specified series of frequencies. The number of frequencies selected depends upon the degree of refinement of the code. Typically, three frequencies may be used to obtain three intensities (decibel level), one for each frequency. The English language has sixty-two phonemes.
30 Thus, each of the sixty-two phonemes will be coded at three selected frequencies. A discussion of voice analysis may be found in the publication Interface Age, issue of May 5, 1977, pages 56-67.

35 Thus, an actor 2, speaking into a microphone 4 transmits phoneme to a voice encoder 6, which digitally encodes the phoneme and transmits the encoded data to a data memory

device 8. This data memory device may be any known type
of electronic digital computer. An example is the model
PDP-11/40 of Digital Equipment Corporation, Maynard, Mas-
sachusetts. An artist may produce a drawing 10 of a par-
5   ticular lip or mouth shape. This drawing 10 may be graph-
ically encoded by the programmer or electronic graphic
scanning device 12. This unit may be of the type described
in United States Patent 3,728,576 and is basically an opti-
cal scanner which encodes the artist's drawing into a
10   graphic digital code for transmission to the computer 8.

The voice encoder 6, previously referred to, is sometimes
known as a speech encoder and is a known piece of equip-
ment. Such a device is sold under the trademark SPEECH
15   LAB and is obtainable from Heuristics, Inc. of Los Altos,
California. The voice encoder 6 is a device which trans-
lates the phoneme into a digital electronic phoneme code.

The artist will draw as many mouth or lip shapes 10 as
20   may be necessary to encode the computer 8 with a complete
phomeme language code and all of the mouth or lip shapes
which may be needed for subsequent graphical reproduction.

Referring now to Fig. 2, there is shown the output or
25   playback mode of the present invention. A keyboard 13 is
used to select a mouth type (male, female, etc.) orienta-
tion front, three-quarter, side, etc. from among the
previously encoded lip or mouth shapes. The keyboard is
of a known type and may be, for example, a DEC LA 36
30   DECWRITER II and/or VT 50 DECSCOPE, products of Digital
Equipment Corporation. The keyboard 13 is connected to
the computer 8 so that the mouth type, mouth orientation,
etc. may be chosen by keying in the desired selection.

35   The actor 2 speaking in the microphone 24 reads a script
or other voice material into the voice encoder 60 which is
similar to the voice encoder 6 previously described. The

voice encoder 60 translates the actor's voice into a
digital electronic voice code. The output of the encoder
60 is transmitted to the computer 8. Under control of the
keyed-in signal from the keyboard 13 and of the encoded
5   output of the voice encoder 60, the data memory device or
computer 8 sends to display device 14 signals corresponding
to the selected graphic code from its memory and also the
phoneme code which most closely matches the actor's encoded
voice. This display device 14 converts the graphic codes
10   and size information into visual information such as the
lip shape 16 shown. The visual images 16 can be recorded
on film or other audio/visual media. For example, visual
images may be enlarged into framed transparencies for over-
lay into compounded frames.
15

Thus, the playback mode of the present arrangement shown in
Fig. 2 allows a simple selection of mouth orientation and
related mouth characteristics to be simply keyed into the
computer which has the various mouth information stored
20   therein. At the same time the voice of the actor 2 may
be encoded to provide an input signal to the computer caus-
ing it to produce a phoneme output most nearly in accord-
ance with the coded signals. As a result, the output from
the computer 8 to the graphic display 14 is controlled by
25   the keyboard input from the keyboard 13 and the output
from the voice encoder 60, the latter of which determines
the lip configuration shown in the graphic display 16.
Thus, if the actor pronounces an "ah" sound into the
microphone 24, the coded input signal to the computer 8
30   will find or select the nearest phoneme code in accordance
with known data comparison techniques. This code will then
be used to provide a predetermined output to display device
14 that will result in an "ah" shaped lip configuration in
display 16.
35

The graphic display device 14 is itself a known item and
may, for example, be a RAMTEK G-100-A color display system,
sold by Ramtek Corporation of Sunnydale, California.

It is possible to overlay directly the constructed visual
image 16 onto an existing film or other audio/visual medium
automatically. In such procedure the original film is
converted by an electronic graphic scanning device, such
5 as is shown at 12 in Fig. 1 and previously described,
into what is known as "pixels". These are electronic
digital codes representing the light intensity at a large
number of points on the screen. The "pixels" are analyzed
by an electronic digital computer by various algorithms
10 to determine the size, orientation and/or location of fea-
tures (in this case the mouth). The pixels in the local
region of the located mouth can be replaced in the elec-
tronic digital computer memory by existing computer instruc-
tions of graphic codes from the sets of phoneme graphic
15 codes stored previously therein and selected by the ar-
rangement shown and described with respect to Fig. 2. The
resulting pixels representing the original frame with
mouth replaced can be sent to an electronic graphic display
device for display and recording.

20
Fig. 3 and Fig. 4 show a modified form of the invention
which may be used for correcting the lip movements in mo-
tion picture film. Fig. 3 shows a motion picture film 20
having a series of frames 22, 24 etc. that include a
25 visual image 25 and a sound track 26. The sound of the
sound track may be a foreign language dubbed in, result-
ing in a sound which does not correspond to the lip move-
ments in the various frames. Accordingly, the film 20 may
be run through a sound projector 28 that embodies a frame
30 counter that sends frame count output pulses over conductor
30 to the digital memory device or computer 8. The sound
projector 28 also projects an image 32 on a suitable
screen. This screen may be a so-called inter-active
graphic tablet. A stylus 34 is used in a known fashion to
35 select the mouth position relative to coordinates on the
graphic tablet 32. The stylus 34 records the position of
the mouth as a digital code and in accordance with known

techniques transmits the information over a conductor 36
for storage into the computer 8. If needed, a keyboard 40
is also utilized whereby data representing a mouth type or
other configuration may be transmitted to the computer 8.

5

An encoder 6a may also be used and is of the type similar
to the encoder 6 previously described. This encoder trans-
mits the digital phoneme into the computer 8. Further-
more, the output sound from the projector as an electrical
10 signal is transmitted over conductor 42 to the encoder 6a,
such electrical signal representing the output sound from
the sound track 26.

Thus, the digital computer 8 has stored therein considerable
15 data in coded form. This data consists of the frame counts,
the mouth position, the phoneme code, and the mouth type.

Turning now to Fig. 4, the playback or output arrangement is
shown. The images 25 of frames 22, 24 etc. are scanned by
20 a conventional optical scanner 50 which sends a digitally
coded image for each frame over conductor 52 to the com-
puter 8. At the same time a pulse advance is supplied over
conductor 54 to advance the frame of the film. The output
signal from the digital computer 8 is sent to a graphic
25 output device 56 which provides a graphic display 58 that
has the new lip movements thereon. Thus, the arrangement
provides for the encoding of the sound from the sound
track 26 and utilizing that data to create a new lip con-
figuration corresponding to the sound of the sound track.
30 The graphic display 58 may be recombined with the sound
in the form of a new film,videotape, or the like.

CLAIMS:

1. A method of creating visual images of lip movements corresponding to speech sounds characterized in providing
5   a coded representation of speech sounds that are associated with lip movements from visual information, which lip movements do not correspond to the speech sounds, transmitting said coded representation to a data memory device (8) for storage therein, storing in said data memory device (8)
10   coded data for creating a graphical representation of new lip movements (58) corresponding to the speech sounds, and transmitting from said data memory device to a graphical output device (56) and under control of the data representing the coded speech, data for the new lip movements cor-
15   responding to said speech sounds.

2. A method according to claim 1 further characterized in applying said new lip movements to a motion picture film to provide new lip movements thereon that are in accord-
20   ance with said speech sounds.

3. A method according to claim 2 further characterized in that the speech sounds are on a sound track (26) on said film.
25

4. A method according to any one of claims 1-3 in which said new lip movements are applied to correct said lip movements that do not correspond to the speech sounds.

30 5. A method according to claim 1 further characterized in extracting the coded speech sounds from a series of frames (22, 24) of audio and said visual information, storing said coded speech sounds in said data memory device on a frame-by-frame basis, optically scanning said series of frames
35   in sequence to provide an encoded graphical image of the visual information and transmitting the encoded graphical image data to said data memory device (8), and transmitting

from said data memory device to said graphical output device (56) the visual information data and the data for the new lip movements.

5   6. A method according to any one of claims 1-5 further characterized in that said coded data for creating a graphical representation of new lip movements comprises a phoneme encoded representation of speech sounds.

10   7. Apparatus for creating visual images of lip movements corresponding to speech sounds characterized in having means (6, 6a) for providing a coded representation of speech sounds that are associated with lip movements from visual information, a data memory device (8), means for
15 transmitting said coded representation to said data memory device (8) for storage therein, said data memory device having stored therein coded data for creating a graphical representation of predetermined lip movements corresponding to the speech sounds, a graphical output device (56),
20 and means for transmitting to said graphical output device and under control of the coded speech representation data, the data for the predetermined lip movements corresponding to the speech sounds.

25   8. Apparatus according to claim 7 further characterized as including means (6a, 28) for extracting the coded speech sounds from a series of frames of audio and said visual information, means (6a) for transmitting said coded speech sounds to said data memory device on a frame-by-frame
30 basis for storage therein on said basis, and means (50) for optically scanning said series of frames in sequence to provide an encoded graphical image of the visual information.
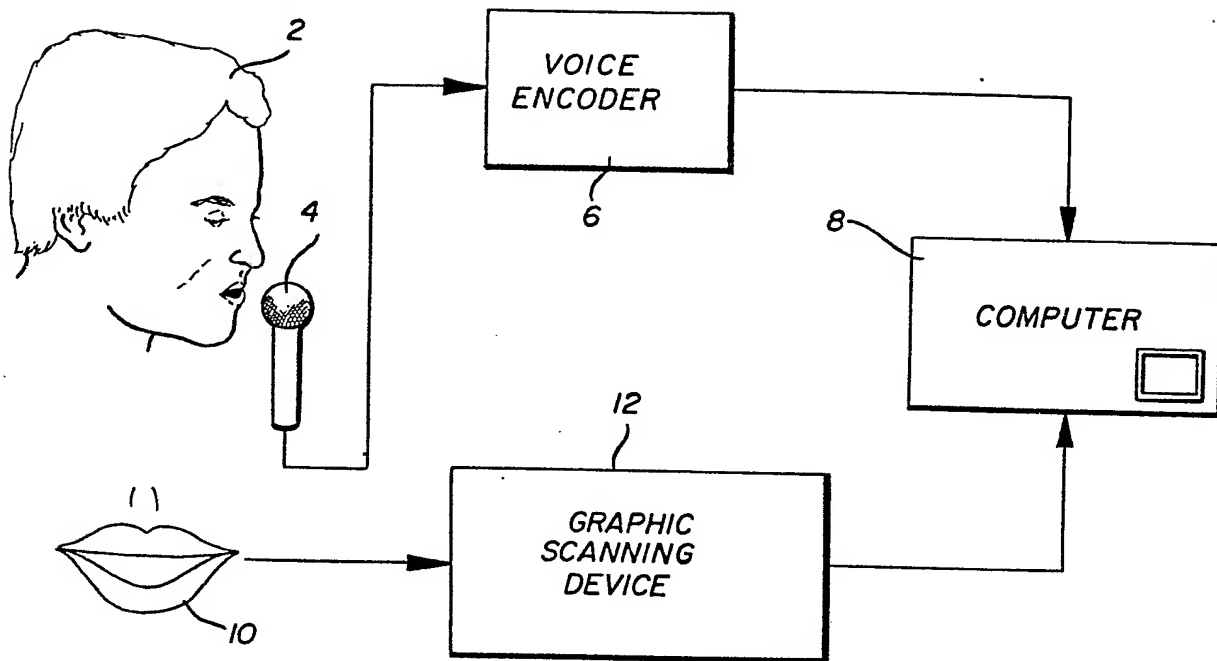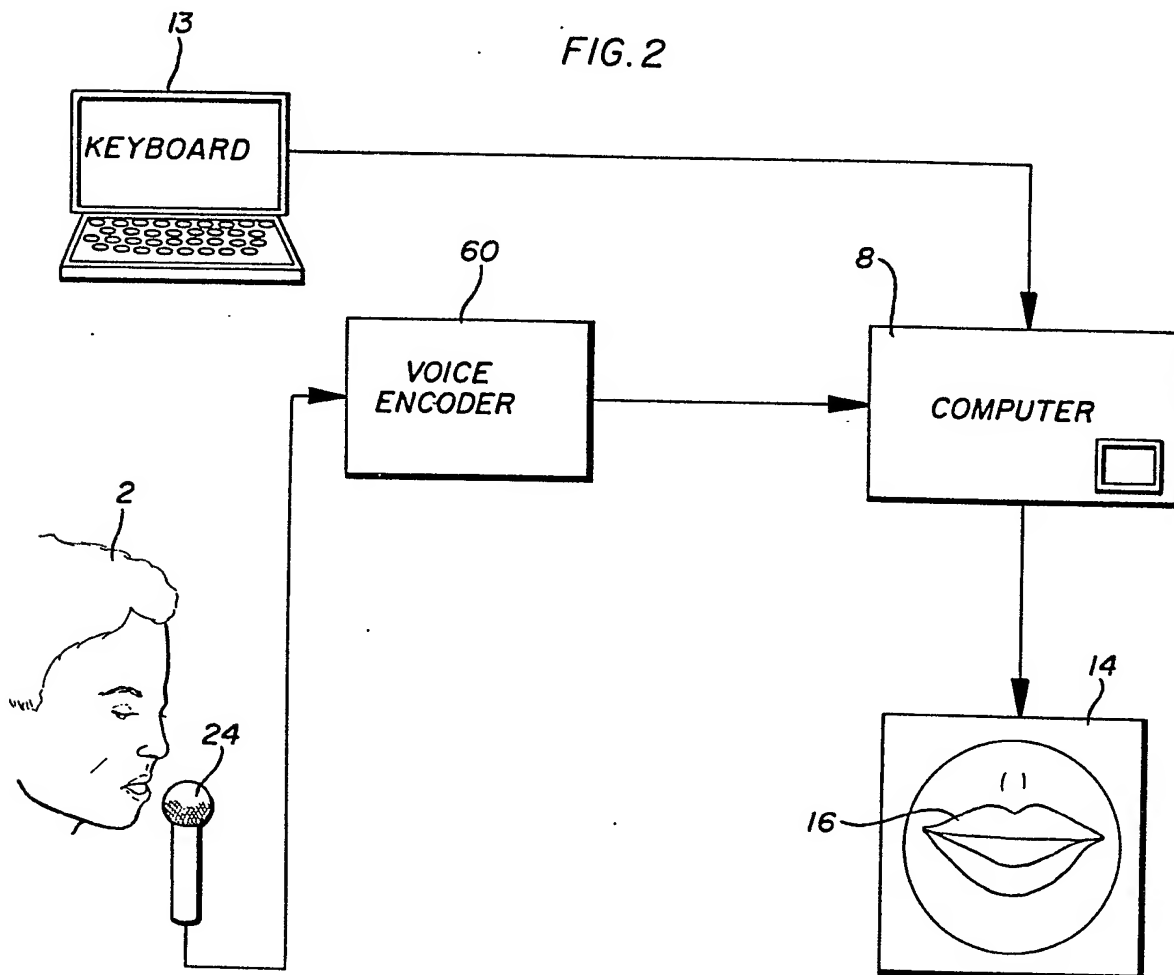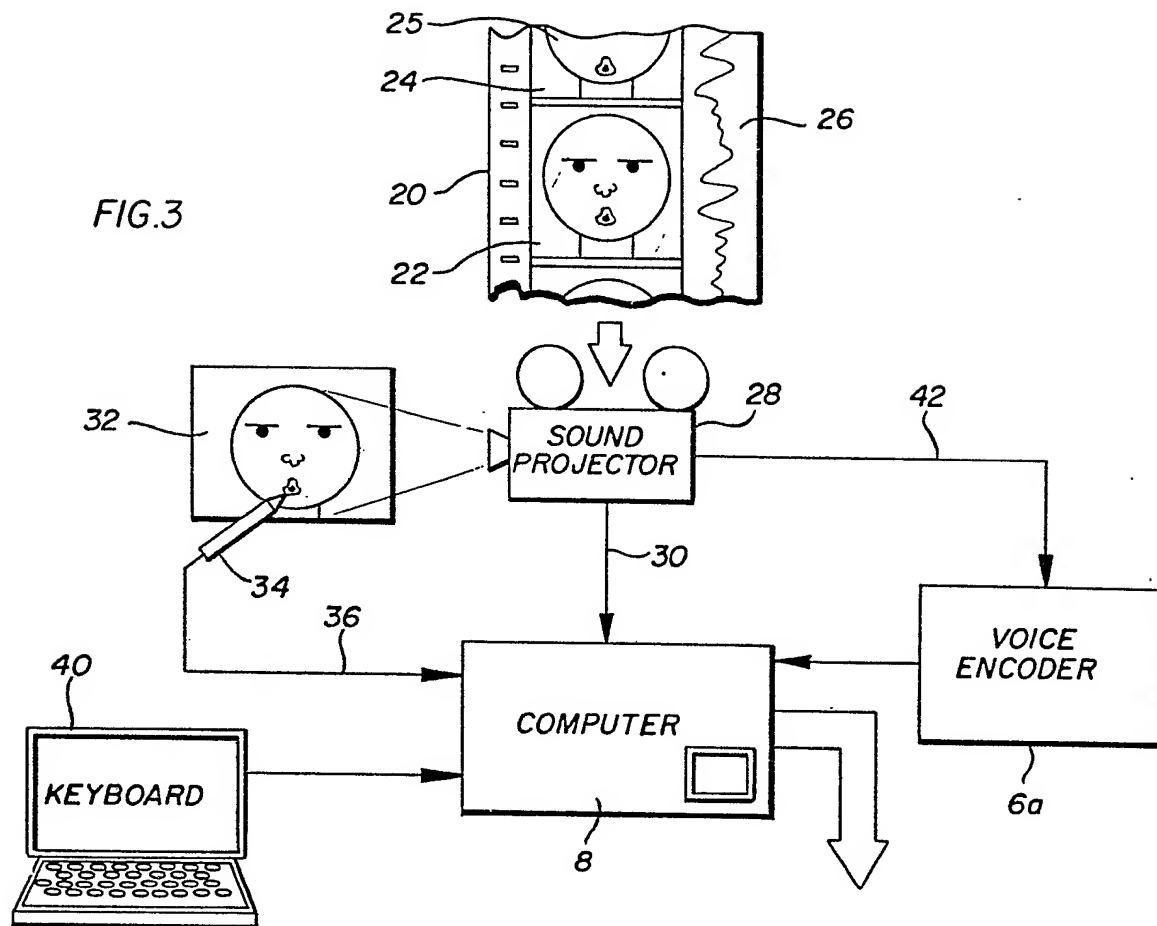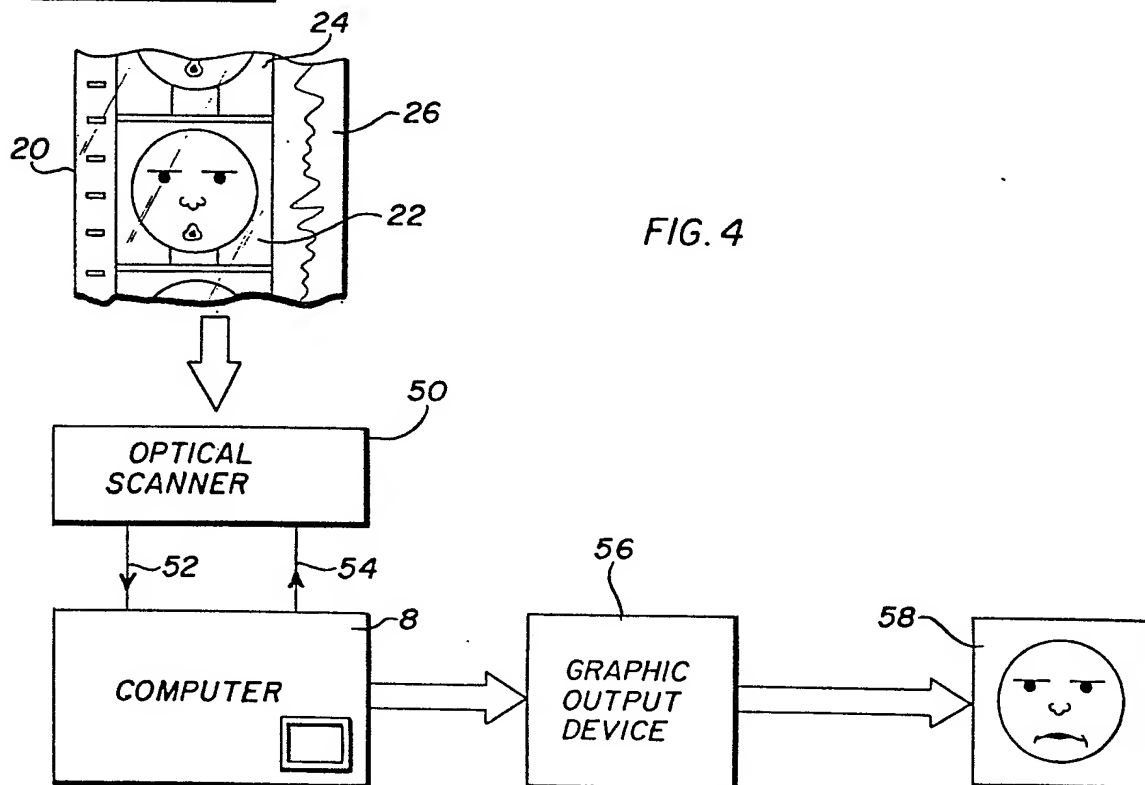
FIG.1

FIG.2

FIG.3

FIG. 4

0056507

European Patent
Office

EUROPEAN SEARCH REPORT

Application number

EP 81 30 0211

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.³)
	BRITISH KINEMATOGRAPHY SOUND AND TELEVISION, vol.53, no.12, December 1971 A. KITCHING:"Computer Animation: Answer or Problem?" pages 436 to 441 * page 439, right-hand column, lines 62 to 72 * --	1-3	G 03 B 15/08 G 09 B 21/00
	US - A - 3 662 374 (HARRISON) * column 1, lines 5 to 29 * --	1	
	US - A - 3 639 691 (BENINGHOF) * claims 5,6 * -----	1,6	

DOCUMENTS CONSIDERED TO BE RELEVANT

TECHNICAL FIELDS SEARCHED (Int. Cl.³)

G 03 B 15/08
        15/10
G 09 B 19/04
-- 19/06
        B21/00
        /06
G 10 L 1/00

CATEGORY OF CITED DOCUMENTS

X: particularly relevant
A: technological background
O: non-written disclosure
P: intermediate document
T: theory or principle underlying the invention
E: conflicting application
D: document cited in the application
L: citation for other reasons
&: member of the same patent family, corresponding document

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	10-09-1981	MEES

EPO Form 1503.1   06.78